Sprung *Journal of Inequalities and Applications* (2019) 2019:4
https://doi.org/10.1186/s13660-018-1953-y

Journal of Inequalities and Applications
a SpringerOpen Journal

RESEARCH **Open Access**

# Upper and lower bounds for the Bregman divergence

Benjamin Sprung[1*]

*Correspondence:
b.sprung@math.uni-goettingen.de
[1]Institute for Numerical and Applied Mathematics, Göttingen, Germany

**Abstract**

In this paper we study upper and lower bounds on the Bregman divergence $\Delta_{\mathcal{F}}^{\xi}(y,x) := \mathcal{F}(y) - \mathcal{F}(x) - \langle \xi, y - x\rangle$ for some convex functional $\mathcal{F}$ on a normed space $\mathcal{X}$, with subgradient $\xi \in \partial\mathcal{F}(x)$. We give a considerably simpler new proof of the inequalities by Xu and Roach for the special case $\mathcal{F}(x) = \|x\|^p$, $p > 1$. The results can be transferred to more general functions as well.

**Keywords:** Bregman divergence; Bregman distance; Uniform smoothness; Uniform convexity; Total convexity

## 1 Introduction

In recent times the Bregman divergence (or Bregman distance) $\Delta_{\mathcal{F}}^{x^*}(y,x)$, introduced by Bregman in [1], has been used as a generalized distance measure in various branches of applied mathematics, for example optimization, inverse problems, statistics and computational mathematics, especially machine learning. To get an overview over the Bregman divergence and its possible applications in optimization and inverse problems we refer to [2–4]. In particular the Bregman divergence has been used for various algorithms in numerical analysis and also for convergence analysis of numerical methods and algorithms.

Especially when doing convergence analysis it is often crucial to have lower and upper bounds on the Bregman divergence in terms of norms. In [5] the authors prove upper and lower bounds for expressions

$$\|x+y\|^p - \|x\|^p - p\langle j_p(x), y\rangle =: \Delta_{\mathcal{F}}^{j_p(x)}(x+y,x), \tag{1}$$

where $j_p : \mathcal{X} \to \mathcal{X}^*$ is a duality mapping, under certain assumptions on the Banach space $\mathcal{X}$. As it turns out that (1) is the Bregman divergence corresponding to the functional $\mathcal{F} = \|\cdot\|^p$ these results have been used since then in many papers working with the Bregman divergence. However, from the proofs of [5] it seems difficult to transfer the results to other functions $\mathcal{F}$. Thus we develop in this work a simple framework to find such bounds and in fact can apply it to give a short new proof of the results from [5] for $\mathcal{F}(x) = \|x\|^p$, $p > 1$.

Our approach is as follows: Proving upper bounds is rather simple if one sufficiently understands the smoothness of $\mathcal{F}$ as the Bregman divergence is basically a linearization error and linearization errors are related to differentiability by definition. In particular we

© The Author(s) 2019. This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

will show that one can obtain upper bounds for the Bregman divergence corresponding to $\mathcal{F} = \phi(\|\cdot\|)$, if $\phi : \mathbb{R} \to \mathbb{R}$ is convex and sufficiently smooth.

Regarding lower bounds we will make use of $\mathcal{F}^*$, the convex conjugate of $\mathcal{F}$. Actually it can be shown that lower bounds for $\Delta_{\mathcal{F}}^{x^*}(y,x)$ correspond to upper bounds for $\Delta_{\mathcal{F}^*}^{x}(y^*,x^*)$. Note that this idea is not at all new. Already in [6, 7] this kind of connection between $\mathcal{F}$ and $\mathcal{F}^*$ was discussed in depth. So again one can just make use of the smoothness of $\mathcal{F}^*$ to conclude lower bounds for $\Delta_{\mathcal{F}}^{x^*}(y,x)$. One might argue that convex conjugates can be rather complicated functions and expecting differentiability is too optimistic. This is true to some extent, but actually reasonable lower bounds on $\Delta_{\mathcal{F}}^{x^*}(y,x)$ already imply differentiability of $\mathcal{F}^*$ at $x^*$ (see [6, Theorem 2.1]). So if one has any hope of finding lower bounds then one might as well work with the convex conjugate.

One reason why our proof is simpler than the proof from [5] is that they did it the other way round. They firstly proved lower bounds with quite some effort and then used the convex conjugate to show upper bounds.

We will focus mainly on asymptotic bounds for $\Delta_{\mathcal{F}}^{x^*}(y,x)$ as $\|x-y\| \to 0$. It is the more interesting case for applications as for example in convergence analysis one will be interested in the Bregman divergence of $x_n$ and $x$, where $x_n \to x$. Also theoretical it is the more challenging case, since for $\|x-y\| \to \infty$ the Bregman divergence $\Delta_{\mathcal{F}}^{x^*}(y,x)$ will mostly depend on the behavior of $\mathcal{F}(y)$ as $y$ tends to infinity and it should be easy to find lower and upper bounds. In particular we will show at the end of the paper, how one can deduce uniform bounds for all $x,y \in \mathcal{X}$ from the asymptotic bounds for the case $\mathcal{F} = \|\cdot\|^p$.

The paper consists of 4 sections. In Sect. 2 we recall some basis notions of convex analysis. In Sect. 3 we define moduli of smoothness and convexity corresponding to a general functional $\mathcal{F}$ and develop some properties of them. Finally, in Sect. 4 we then use the theory from Sect. 3 on the functional $\mathcal{F} = \frac{1}{p}\|\cdot\|^p$ for $p > 1$ and find lower and upper bounds for the corresponding Bregman divergence given by the smoothness, respectively, the convexity of the space $\mathcal{X}$ as shown in [5].

## 2 Tools from convex analysis

In this work $\mathcal{X}$ will always be a real Banach space, with $\dim \mathcal{X} \geq 2$, $\mathcal{X}^*$ denotes its dual space, $S_{\mathcal{X}} = \{x \in \mathcal{X} : \|x\| = 1\}$ the unit sphere and $\mathcal{F} : \mathcal{X} \to \overline{\mathbb{R}} := \mathbb{R} \cup \{\infty\}$ some function. We will need some basic concepts from convex analysis, so we shortly recall them in this chapter.

$x^* \in \mathcal{X}^*$ is called a *subgradient* of a convex function $\mathcal{F} : \mathcal{X} \to \overline{\mathbb{R}}$ at $x \in \mathcal{X}$ if $\mathcal{F}(x)$ is finite and

$$\mathcal{F}(y) \geq \mathcal{F}(x) + \langle x^*, y-x \rangle, \tag{2}$$

for all $y \in \mathcal{X}$. The set of all subgradients of $\mathcal{F}$ at $x$ is called the *subdifferential* of $\mathcal{F}$ at $x$ and denoted by $\partial\mathcal{F}(x)$. The *convex conjugate* $\mathcal{F}^* : \mathcal{X}^* \to \overline{\mathbb{R}}$ of $\mathcal{F}$ is defined by

$$\mathcal{F}^*(x^*) = \sup_{x \in \mathcal{X}} \bigl[\langle x^*, x \rangle - \mathcal{F}(x)\bigr].$$

From these two definitions one can directly conclude the following *generalized Young (in)equality*. For all $x \in \mathcal{X}$, $x^* \in \mathcal{X}^*$ we have

$$\mathcal{F}(x) + \mathcal{F}^*(x^*) \geq \langle x^*, x \rangle. \tag{3}$$

Equality holds true if and only if $x^* \in \partial\mathcal{F}(x)$. Further we have

$$\mathcal{F}(x) \geq \mathcal{F}^{**}(x) := \sup_{x^* \in \mathcal{X}^*} \left[\langle x^*, x\rangle - \mathcal{F}^*(x^*)\right], \quad x \in X, \tag{4}$$

where equality holds if and only if $\mathcal{F}$ is convex and lower-semicontinuous. Finally, we define the object of interest of this work. For $\mathcal{F}(x) < \infty$ and $x^* \in \partial\mathcal{F}(x)$ the *Bregman divergence* $\Delta_{\mathcal{F}}^{x^*}(y,x)$ is given by

$$\Delta_{\mathcal{F}}^{x^*}(y,x) = \mathcal{F}(y) - \mathcal{F}(x) - \langle x^*, y - x\rangle \geq 0,$$

for all $y \in \mathcal{X}$. We will be especially interested in functionals $\mathcal{F}(x) = \frac{1}{p}\|x\|^p$ for some $p \geq 1$ and need to understand their subdifferentials, so finally we have the following. For some $p \geq 1$ the set-valued mapping $J_p : \mathcal{X} \to 2^{\mathcal{X}^*}$ given by

$$J_p(x) = \left\{x^* \in \mathcal{X}^* : \langle x^*, x\rangle = \|x^*\|\,\|x\|, \|x^*\| = \|x\|^{p-1}\right\}$$

is called the *duality mapping* with respect to $p$ of $\mathcal{X}$. The sets $J_p(x)$ are always non-empty. A mapping $j_p : \mathcal{X} \to \mathcal{X}^*$ is called *selection* of $J_p$ if $j_p(x) \in J_p(x)$ for all $x \in \mathcal{X}$. If $\mathcal{F}(x) = \frac{1}{p}\|x\|^p$, then the theorem of Asplund [8, Theorem 1] yields

$$\partial\mathcal{F}(x) = J_p(x).$$

## 3 Moduli of smoothness and convexity

Finding upper bounds for (1) is related to the smoothness of the norm of $\mathcal{X}$ whereas lower bounds are related to convexity. Thus it is necessary to understand the moduli of smoothness and convexity of the space $\mathcal{X}$ and we shortly recall their definitions (see e.g. [9]):

**Definition 3.1** The modulus of convexity $\delta_{\mathcal{X}} : [0,2] \to [0,1]$ of the space $\mathcal{X}$ is defined by

$$\delta_{\mathcal{X}}(\varepsilon) := \inf\left\{1 - \|y + \tilde{y}\|/2 : y, \tilde{y} \in S_{\mathcal{X}}, \|y - \tilde{y}\| = \varepsilon\right\}.$$

The modulus of smoothness $\rho_{\mathcal{X}} : [0,\infty) \to [0,\infty)$ of $\mathcal{X}$ is defined by

$$\rho_{\mathcal{X}}(\tau) := \sup\left\{\left(\|x + \tau y\| + \|x - \tau y\|\right)/2 - 1 : x, y \in S_{\mathcal{X}}\right\}.$$

The space $\mathcal{X}$ is called *uniformly convex* if $\delta_{\mathcal{X}}(\varepsilon) > 0$ for every $\varepsilon > 0$. It is called *uniformly smooth* if $\lim_{\tau\to 0} \rho_{\mathcal{X}}(\tau)/\tau = 0$. The space $\mathcal{X}$ is called *r-convex* (or convex of power type $r$) if there exists a constant $K > 0$ such that $\delta_{\mathcal{X}}(\varepsilon) \geq K\varepsilon^r$ for all $\varepsilon \in [0,2]$. Similarly, it is called $s$-smooth (or smooth of power type $s$) if $\rho_{\mathcal{X}}(\tau) \leq K\tau^s$ for all $\tau > 0$.

These two moduli have a well-developed theory, which has been known in the literature for a long time and we will not discuss all their properties. However, for our proofs we will need the following properties.

**Lemma 3.2**

1. *We have for $\tau_1 \leq \tau_2$ that $\rho_{\mathcal{X}}(\tau_1)/\tau_1 \leq \rho_{\mathcal{X}}(\tau_2)/\tau_2$.*

2. *We see for all* $\overline{\tau} > 0$ *that there exists a constant* $C_{\overline{\tau}}$ *such that for all Banach spaces* $\mathcal{X}$ *we have*

$$\rho_{\mathcal{X}}(\tau) \geq \left(1+\tau^2\right)^{\frac{1}{2}} - 1 \geq C_{\overline{\tau}}\tau^2, \quad \tau \leq \overline{\tau}.$$

3. *If* $\delta_{\mathcal{X}}$ *is extended by* $\infty$ *on* $\mathbb{R} \setminus [0,2]$ *then* $(2\delta_{\mathcal{X}})^* = 2\rho_{\mathcal{X}^*}$.
4. *There exists a convex function* $f$ *such that* $\delta_{\mathcal{X}}(\tau/2) \leq f(\tau) \leq \delta_{\mathcal{X}}(\tau)$. *In particular we have* $\delta_{\mathcal{X}}^{**}(\tau) \geq \delta_{\mathcal{X}}(\tau/2)$.

*Proof* All statements follow from [9, Ch. 1.e]. The function $f$ in the last statement can be chosen as the Orlicz function $f(\tau) := (\rho_{\mathcal{X}^*})(\tau/2)$ by [9, Lemmata 1.e.6, 1.e.7]. □

For our purposes it will be more natural to introduce new definitions of the moduli of smoothness and convexity related to functionals instead of spaces.

**Definition 3.3** Let $\mathcal{F}\colon \mathcal{X} \to \overline{\mathbb{R}}$ be some arbitrary function, $x \in \mathcal{X}$, $\mathcal{F}(x) < \infty$ and $\xi \in \mathcal{X}^*$. Define the *linearization error functional* $\Delta_{\mathcal{F}}^{\xi}(y,x)$ by

$$\Delta_{\mathcal{F}}^{\xi}(y,x) = \mathcal{F}(y) - \mathcal{F}(x) - \langle \xi, y-x \rangle.$$

The modulus of smoothness $\rho_{\mathcal{F},x}^{\xi}\colon [0,\infty) \to [0,\infty]$ of $\mathcal{F}$ in $x$ with respect to $\xi$ is defined by

$$\rho_{\mathcal{F},x}^{\xi}(\tau) := \sup_{y \in S_{\mathcal{X}}} \left|\mathcal{F}(x+\tau y) - \mathcal{F}(x) - \langle \xi, \tau y \rangle\right| = \sup_{\|x-y\|=\tau} \left|\Delta_{\mathcal{F}}^{\xi}(y,x)\right|.$$

The modulus of convexity $\delta_{\mathcal{F},x}^{\xi}\colon [0,\infty) \to [0,\infty]$ of $\mathcal{F}$ in $x$ with respect to $\xi$ is defined by

$$\delta_{\mathcal{F},x}^{\xi}(\tau) := \inf_{\|x-y\|=\tau} \left|\Delta_{\mathcal{F}}^{\xi}(y,x)\right|.$$

$\mathcal{F}$ is called *r-convex* (or convex of power type $r$) in $x$ (w.r.t. $\xi$) if there exist $K, \overline{\tau} > 0$ such that $\delta_{\mathcal{F},x}^{\xi}(\tau) \geq K\tau^r$ for all $0 < \tau \leq \overline{\tau}$. Similarly, it is called $s$-smooth (or smooth of power type $s$) in $x$ (w.r.t. $\xi$) if $\rho_{\mathcal{F},x}^{\xi}(\tau) \leq K\tau^s$ for all $0 < \tau \leq \overline{\tau}$.

The quantities $\rho_{\mathcal{F},x}^{\xi}$, $\delta_{\mathcal{F},x}^{\xi}$ give us a reformulation of our basic problem: We want to find upper bounds for $\rho_{\mathcal{F},x}^{\xi}(\tau)$ and lower bounds for $\delta_{\mathcal{F},x}^{\xi}(\tau)$. Before we show some properties of these functions we should state some simple facts for their interpretation.

*Remark* 3.4 We will mostly consider convex functions $\mathcal{F}$ with $\xi \in \partial\mathcal{F}(x)$ so that the linearization error functional is a Bregman divergence and one can neglect the absolute value.

$\mathcal{F}$ is Fréchet-differentiable in $x$ if and only if there exists $\xi \in \mathcal{X}^*$, such that $\rho_{\mathcal{F},x}^{\xi}(\tau)/\tau \to 0$ as $\tau \to 0$. $\mathcal{F}$ being $s$-smooth in $x$, with $s \in (1,2]$ then can be seen as a stronger form of differentiability, comparable to fractional derivatives; however, $\mathcal{F}$ being 2-smooth is not equivalent to twice differentiability but rather to the notion of strong smoothness.

If there exists a selection $j : \mathcal{X} \to \mathcal{X}^*$ of the subdifferential of $\mathcal{F}$, i.e. for every $x$ exists $j(x) \in \partial\mathcal{F}(x)$, then this implies already that $\mathcal{F}$ is convex. $\delta_{\mathcal{F},x}^{j(x)}(\tau) > 0$ for all $x, \tau$ implies strict convexity. As before $r$-convexity is an even stronger notion of convexity and 2-convexity is connected to strong convexity. In [10] the *modulus of local (or total) convexity of* $\mathcal{F}$,

$\nu_{\mathcal{F}}(x,\tau)$, was introduced and is basically given by $\delta^{\xi}_{\mathcal{F},x}(\tau)$ just that $\langle \xi, y-x\rangle$ is replaced by the right hand side derivative of $\mathcal{F}$ at $x$ in direction $y-x$. If $\mathcal{F}$ is convex and Gâteaux-differentiable then $\nu_{\mathcal{F}}(x,\tau)$ coincides with $\delta^{\xi}_{\mathcal{F},x}(\tau)$, where $\xi = \mathcal{F}'(x)$. The modulus of total convexity has been studied in several papers, see e.g. [11]. There exist further definitions of moduli of convexity and smoothness related to functions (e.g. [4, 12]), but giving a complete overview over all such definitions goes beyond the scope of this work.

It turns out that for functionals $\mathcal{F}$ that originate from the norm of $\mathcal{X}$ the moduli of the space and of the functions are closely related.

**Proposition 3.5** *Let $\mathcal{F} = \|\cdot\|_{\mathcal{X}}$ and for all $x \in \mathcal{X}$ let $\xi_x \in \partial\mathcal{F}(x)$ be arbitrary. We have*

$$\rho \le \sup_{x\in S_{\mathcal{X}}} \rho^{\xi_x}_{\mathcal{F},x} \le 2\rho. \tag{5}$$

*Proof* If we replace $y$ by $-y$ in the definition of $\rho^{\xi_x}_{\mathcal{F},x}$ we see

$$\begin{aligned} 2\rho(\tau) &= \sup\bigl\{\mathcal{F}(x+\tau y) + \mathcal{F}(x-\tau y) - 2\mathcal{F}(x) + \langle \xi_x, y-y\rangle : x,y \in S_{\mathcal{X}}\bigr\} \\ &\le 2 \sup_{x\in S_{\mathcal{X}}} \rho^{\xi_x}_{\mathcal{F},x}(\tau) \end{aligned}$$

and for all $x,y \in S_{\mathcal{X}}$ we have by the definition of the subdifferential and as $\mathcal{F}(x) = \|x\|_{\mathcal{X}} = 1$

$$\mathcal{F}(x+\tau y) - \mathcal{F}(x) - \langle \xi_x, \tau y\rangle \le \mathcal{F}(x+\tau y) + \mathcal{F}(x-\tau y) - 2 \le 2\rho(\tau). \qquad \square$$

So this already gives us an upper bound for $\rho^{\xi}_{\|\cdot\|_{\mathcal{X}},x}(\tau)$ if $x \in S_{\mathcal{X}}$, $\xi \in \partial\mathcal{F}(x)$. For generalizing this to all $x \in \mathcal{X}$ we use the following.

**Proposition 3.6** *If the functional $\mathcal{F}$ is positively $q$-homogeneous then we have, for all $x \in \mathcal{X}\setminus\{0\}$, $\xi \in \mathcal{X}^*$,*

$$\|x\|^q \delta^{\xi/\|x\|^{q-1}}_{\mathcal{F},x/\|x\|}\left(\frac{\|x-y\|}{\|x\|}\right) \le \bigl|\Delta^{\xi}_{\mathcal{F}}(y,x)\bigr| \le \|x\|^q \rho^{\xi/\|x\|^{q-1}}_{\mathcal{F},x/\|x\|}\left(\frac{\|x-y\|}{\|x\|}\right)$$

*and $\xi/\|x\|^{q-1} \in \partial\mathcal{F}(x/\|x\|)$ if and only if $\xi \in \partial\mathcal{F}(x)$.*

*Proof* If $\mathcal{F}$ is positively $q$-homogeneous we have

$$\bigl|\Delta^{\xi}_{\mathcal{F}}(y,x)\bigr| = \|x\|^q \left|\Delta^{\xi/\|x\|^{q-1}}_{\mathcal{F}}\left(\frac{y}{\|x\|}, \frac{x}{\|x\|}\right)\right|,$$

so that the first claim follows from Definition 3.3. The second claim follows from multiplying (2) either by $\|x\|^q$ or $\|x\|^{-q}$. $\square$

For convex functions $\mathcal{F}$ one can show that both moduli are nondecreasing.

**Proposition 3.7** *Let $\mathcal{F}$ be convex, $x \in \mathcal{X}$ and $\xi \in \partial\mathcal{F}(x)$. Then for $\lambda \ge 1$ one has*

$$\rho^{\xi}_{\mathcal{F},x}(\lambda\tau) \ge \lambda\rho^{\xi}_{\mathcal{F},x}(\tau), \qquad \delta^{\xi}_{\mathcal{F},x}(\lambda\tau) \ge \lambda\delta^{\xi}_{\mathcal{F},x}(\tau).$$

*In particular $\delta^{\xi}_{\mathcal{F},x}$, $\rho^{\xi}_{\mathcal{F},x}$ are nondecreasing.*

*Proof* The idea is the same, as in [10, Sect. 2.4]. Let $\lambda \geq 1$. For all $y \in \mathcal{X}$, $\|y - x\| = \tau$ one can define $y_\lambda = \lambda y + (1-\lambda)x$, so $\|y_\lambda - x\| = \lambda\tau$. As $\lambda \geq 1$ we get by the convexity of $\mathcal{F}$

$$\mathcal{F}\big(\lambda y + (1-\lambda)x\big) \geq \lambda\mathcal{F}(y) + (1-\lambda)\mathcal{F}(x)$$

(multiply by $\lambda^{-1}$ and replace $y = \lambda^{-1}y + (1-\lambda^{-1})x$ to see the equivalence to the usual definition of convexity) and thus

$$\frac{1}{\lambda}\Delta^{\xi}_{\mathcal{F}}(y_\lambda, x) = \frac{1}{\lambda}\big(\mathcal{F}\big(\lambda y + (1-\lambda)x\big) - \mathcal{F}(x)\big) - \langle \xi, y - x\rangle \geq \Delta^{\xi}_{\mathcal{F}}(y, x).$$

So for all $y \in \mathcal{X}$, $\|y - x\| = \tau$ we find

$$\Delta^{\xi}_{\mathcal{F}}(y, x) \leq \frac{1}{\lambda}\rho^{\xi}_{\mathcal{F},x}(\lambda\tau),$$

which gives the first inequality. Similarly for all $y \in \mathcal{X}$, $\|y - x\| = \lambda\tau$ one can define $\tilde{y}_\lambda = \frac{1}{\lambda}y + (1 - \frac{1}{\lambda})x$, then $\|\tilde{y}_\lambda - x\| = \tau$ and again convexity of $\mathcal{F}$ can be used to show $\Delta^{\xi}_{\mathcal{F}}(y, x) \geq \lambda\Delta^{\xi}_{\mathcal{F}}(\tilde{y}_\lambda, x)$, which yields the other inequality. □

We also have a chain rule.

**Proposition 3.8** *Let $f\colon \mathbb{R} \to \mathbb{R}$ and $x \in \mathcal{X}$, $\xi \in \mathcal{X}^*$, $t \in \mathbb{R}$ be such that $\rho^t_{f,\mathcal{F}(x)}$ is nondecreasing. Then for all $\tau \geq 0$ we have*

$$\rho^{t\xi}_{f\circ\mathcal{F},x}(\tau) \leq |t|\rho^{\xi}_{\mathcal{F},x}(\tau) + \rho^t_{f,\mathcal{F}(x)}\big(\|\xi\|\tau + \rho^{\xi}_{\mathcal{F},x}(\tau)\big).$$

*Proof* Let $s = \mathcal{F}(x)$ and define functions $R$, $r$ by

$$\mathcal{F}(x+y) - \mathcal{F}(x) = \langle \xi, y\rangle + R(y) \quad \forall y \in \mathcal{X},$$
$$f(s+h) - f(s) = th + r(h) \quad \forall h \in \mathbb{R}.$$

Then we have for $\tau > 0$ and $y \in S_{\mathcal{X}}$

$$\begin{aligned} f\circ\mathcal{F}(x+\tau y) - f\circ\mathcal{F}(x) &= t\big(\langle \xi, \tau y\rangle + R(y)\big) + r\big(\langle \xi, \tau y\rangle + R(\tau y)\big) \\ &= \langle t\xi, \tau y\rangle + tR(\tau y) + r\big(\langle \xi, \tau y\rangle + R(\tau y)\big). \end{aligned}$$

Now the claim follows from $R(\tau y) \leq \rho^{\xi}_{\mathcal{F},x}(\tau)$ and $r(h) \leq \rho^t_{f,\mathcal{F}(x)}(|h|)$ together with the assumption that $\rho^t_{f,\mathcal{F}(x)}$ is a nondecreasing function. □

Propositions 3.5, 3.7 and 3.8 are already sufficient to find upper bounds on $\rho^{\xi}_{\mathcal{F},x}$ for $\mathcal{F} = f(\|x\|_{\mathcal{X}})$ if $f$ is convex and if we sufficiently understand the smoothness of $f$ and of the space $\mathcal{X}$. Regarding lower bounds the following proposition will be our key instrument.

**Proposition 3.9** *Let $\mathcal{F}$ convex and $x$ be such there exists $\xi \in \partial\mathcal{F}(x)$. We have*

$$\big(\delta^{\xi}_{\mathcal{F},x}\big)^* = \rho^{x}_{\mathcal{F}^*,\xi}. \tag{6}$$

*Further we see that $\mathcal{F}$ is $p$-convex in $x$ w.r.t. $\xi$ if and only if $\mathcal{F}^*$ is $p'$-smooth in $\xi$ w.r.t. $x$.*

*Proof* We have

$$\begin{aligned}\rho^x_{\mathcal{F}^*,\xi}(\tau) &= \sup_{y^*\in S_{\mathcal{X}^*}}\left[\mathcal{F}^*\left(\xi+\tau y^*\right)-\mathcal{F}^*(\xi)-\left\langle\tau y^*,x\right\rangle\right]\\ &= \sup_{y^*\in S_{\mathcal{X}^*}}\sup_{y\in\mathcal{X}}\left[\left\langle\xi+\tau y^*,y\right\rangle-\mathcal{F}(y)-\mathcal{F}^*(\xi)-\left\langle\tau y^*,x\right\rangle\right]\\ &= \sup_{y\in\mathcal{X}}\left[\langle\xi,y\rangle-\mathcal{F}(y)-\mathcal{F}^*(\xi)+\tau\|y-x\|\right].\end{aligned}$$

By Young's equality (3) we then have

$$\begin{aligned}\rho^x_{\mathcal{F}^*,\xi}(\tau) &= \sup_{y\in\mathcal{X}}\left[\mathcal{F}(x)-\mathcal{F}(y)+\langle\xi,y-x\rangle+\tau\|y-x\|\right]\\ &= \sup_{\varepsilon\in\mathbb{R}^+_0}\sup_{y\in\mathcal{X},\|y-x\|=\varepsilon}\left[\varepsilon\tau-\Delta^\xi_{\mathcal{F}}(y,x)\right]=\left(\delta^\xi_{\mathcal{F},x}\right)^*(\tau).\end{aligned}$$

The second statement follows from (6), which gives

$$\rho^x_{\mathcal{F}^*,\xi}=\left(\delta^\xi_{\mathcal{F},x}\right)^*,\qquad \delta^\xi_{\mathcal{F},x}\geq\left(\delta^\xi_{\mathcal{F},x}\right)^{**}=\left(\rho^x_{\mathcal{F}^*,\xi}\right)^*$$

and the fact that by Proposition 3.7 we have for $\tau>\overline{\tau}$ $\rho^\xi_{\mathcal{F},x}(\tau)\geq\tau\rho^\xi_{\mathcal{F},x}(\overline{\tau})/\overline{\tau}$, $\delta^\xi_{\mathcal{F},x}(\tau)\geq\tau\delta^\xi_{\mathcal{F},x}(\overline{\tau})/\overline{\tau}$, so that in particular

$$\begin{aligned}\left(\delta^\xi_{\mathcal{F},x}\right)^*\left(\tau^*\right) &= \sup_{0\leq\tau\leq\overline{\tau}}\left[\tau^*\tau-\delta^\xi_{\mathcal{F},x}(\tau)\right],\quad \text{for } \tau^*\leq\delta^\xi_{\mathcal{F},x}(\overline{\tau})/\overline{\tau},\\ \left(\rho^\xi_{\mathcal{F},x}\right)^*\left(\tau^*\right) &= \sup_{0\leq\tau\leq\overline{\tau}}\left[\tau^*\tau-\rho^\xi_{\mathcal{F},x}(\tau)\right],\quad \text{for } \tau^*\leq\rho^\xi_{\mathcal{F},x}(\overline{\tau})/\overline{\tau}.\end{aligned}$$

Thus one can just put in the corresponding lower or upper bound and calculate the maximum, which completes the proof. □

## 4 Application to norm powers

In this section we will consider $\mathcal{F}=\frac{1}{p}\|\cdot\|^p$ for some $p>1$ and use the theory from the last chapter to reproduce the main results from [5]. Note that in the light of Proposition 3.6 it is sufficient to understand $\delta^{j_p(x)}_{\mathcal{F},x}$ and $\rho^{j_p(x)}_{\mathcal{F},x}$ for $x\in S_{\mathcal{X}}$.

**Theorem 4.1** *For some fixed $p>1$ let $\mathcal{F}=\frac{1}{p}\|\cdot\|^p$.*

1. *For all $\overline{\tau}>0$ exists a constant $C_{\overline{\tau},p}>0$, such that for $x\in S_{\mathcal{X}}$ and $\tau\leq\overline{\tau}$ we have*

$$\rho^{j_p(x)}_{\mathcal{F},x}(\tau)\leq C_{\overline{\tau},p}\rho_{\mathcal{X}}(\tau).$$

2. *If we have for $\overline{\tau}>0$, $\tau\leq\overline{\tau}$ and all $x\in S_{\mathcal{X}}$*

$$\rho^{j_p(x)}_{\mathcal{F},x}(\tau)\leq\phi(\tau),$$

*then*

$$\rho_{\mathcal{X}}(\tau)\leq p^{1/p-1}\phi(\tau)+C_{\overline{\tau}}\tau^2,$$

*for* $\tau \le \overline{\tau}$. *In particular if* $\phi\colon \mathbb{R}^+ \to \mathbb{R}^+$ *fulfills* $\lim_{\tau\to 0}\phi(\tau)/\tau = 0$, *then* $\mathcal{X}$ *is uniformly smooth.*

3. *Let* $\frac{1}{p} + \frac{1}{p'} = 1$. *For all* $x \in S_{\mathcal{X}}$, $\overline{\tau} > 0$ *we have*

$$\delta^{j_p(x)}_{\mathcal{F},x}(\tau) \ge C_{\overline{\tau},p'}\delta_{\mathcal{X}}(\tau/C_{\overline{\tau},p'}), \quad \tau \le \frac{p'-1}{2}\overline{\tau},$$

*where* $C_{\overline{\tau},p'}$ *is the constant from* 1.

4. *If there exists* $\overline{\tau} > 0$ *such that we have for all* $x \in S_{\mathcal{X}}$ *and* $\tau \le \overline{\tau}$

$$\delta^{j_p(x)}_{\mathcal{F},x}(\tau) \ge \phi(\tau),$$

*where* $\phi\colon \mathbb{R}^+ \to \mathbb{R}^+$ *is nondecreasing and* $\phi(\tau) > 0$ *for* $\tau > 0$, *then* $\mathcal{X}$ *is uniformly convex.*

*Proof Claim* 1: Note that $\mathcal{F} = f \circ \|\cdot\|$, with $f(t) = \frac{1}{p}t^p$, which is convex, thus $\rho^1_{f,\mathcal{F}(x)}$ is nondecreasing by Proposition 3.7, so Proposition 3.8 gives

$$\rho^{j_p(x)}_{\mathcal{F},x}(\tau) \le \rho^{j_p(x)}_{\|\cdot\|,x}(\tau) + \rho^1_{f,\mathcal{F}(x)}\bigl(\tau + \rho^{j_p(x)}_{\|\cdot\|,x}(\tau)\bigr).$$

We have by Taylor's theorem

$$\rho^1_{f,1}(\tau) = \sup_{\sigma\in\{-1,+1\}} \frac{p-1}{2}\tau^2 + r(\sigma\tau)\tau^2 \le C\tau^2, \quad \text{for } \tau \le 3\overline{\tau},$$

where the inequality holds for a constant depending on $p$ and $\overline{\tau}$ as $\rho^1_{f,1}$ is always finite and so is the remainder $r$. We have $j_p(x) \in \partial\|\cdot\|(x)$ for $x \in S_{\mathcal{X}}$, so by Proposition 3.5 we have $\rho^{j_p(x)}_{\|\cdot\|,x}(\tau) \le 2\rho_{\mathcal{X}}(\tau)$ and one can easily see that $\rho_{\mathcal{X}}(\tau) \le \tau$. So we have

$$\rho^{j_p(x)}_{\mathcal{F},x}(\tau) \le 2\rho_{\mathcal{X}}(\tau) + 9C\tau^2 \le (2 + 9C/C_{\overline{\tau}})\rho_{\mathcal{X}}(\tau), \quad \tau \le \overline{\tau},$$

where the second inequality follows from item 2 of Lemma 3.2.

*Claim* 2: Note that $\|\cdot\| = f^{-1} \circ \mathcal{F}$ and $f^{-1}(t) = (pt)^{\frac{1}{p}}$ is concave, thus $-f^{-1}$ is convex and it is differentiable, so $-1 \in \partial(-f^{-1})(\frac{1}{p})$ and by Proposition 3.7 $\rho^1_{f^{-1},1/p} = \rho^{-1}_{-f^{-1},1/p}$ is nondecreasing. Then Proposition 3.8 gives for all $x \in S_{\mathcal{X}}$

$$\rho^{j_p(x)}_{\|\cdot\|,x}(\tau) \le \rho^{j_p(x)}_{\mathcal{F},x}(\tau) + \rho^1_{f^{-1},1/p}\bigl(\tau + \rho^{j_p(x)}_{\mathcal{F},x}(\tau)\bigr) \le \phi(\tau) + C_{\overline{\tau}}\tau^2,$$

where the second inequality follows by Taylor's theorem as above and the fact that by Claim 1 we always have $\rho^{j_p(x)}_{\mathcal{F},x}(\tau) \le C\tau$ for some $C > 0$. Thus Proposition 3.5 yields the assertion.

*Claim* 3: First of all note that $\mathcal{F}^*(t) = \frac{1}{p'}t^{p'}$, with $\frac{1}{p} + \frac{1}{p'} = 1$. We have

$$\delta^{j_p(x)}_{\mathcal{F},x}(\tau) \ge \bigl(\delta^{j_p(x)}_{\mathcal{F},x}\bigr)^{**}(\tau) = \bigl(\rho^x_{\mathcal{F}^*,j_p(x)}\bigr)^*(\tau) = \sup_{r\ge 0}\bigl[\tau r - \rho^x_{\mathcal{F}^*,j_p(x)}(r)\bigr].$$

By Claim 1 we have for all $x \in S_{\mathcal{X}}$ $\rho^x_{\mathcal{F}^*,j_p(x)}(r) \le C_{\overline{\tau},p'}\rho_{\mathcal{X}^*}(r)$ for all $0 < r < \overline{\tau}$. We are only interested in the case $\tau \to 0$ so let $\tau \le C_{\overline{\tau},p'}\rho_{\mathcal{X}^*}(\overline{\tau})/\overline{\tau}$, where $\rho_{\mathcal{X}^*}(\overline{\tau})/\overline{\tau} > 0$ by Lemma 3.2, 2.

Then by Lemma 3.2, 1. we have $\tau r \le C_{\overline{\tau},p'}\rho_{\mathcal{X}^*}(r)$ for $r \ge \overline{\tau}$ and thus find

$$\sup_{0\le r}\left[\tau r - \rho^{x}_{\mathcal{F}^*,j_p(x)}(r)\right] \ge \sup_{0\le r\le\overline{\tau}}\left[\tau r - C_{\overline{\tau},p'}\rho_{\mathcal{X}^*}(r)\right] = (C\rho_{\mathcal{X}^*})^*(\tau).$$

So we have by Lemmas 3.2, 3 and 4

$$\delta^{j_p(x)}_{\mathcal{F},x}(\tau) \ge (C_{\overline{\tau},p'}\rho_{\mathcal{X}})^*(\tau) = \frac{C_{\overline{\tau},p'}}{2}(2\delta_{\mathcal{X}})^{**}\left(\frac{2\tau}{C_{\overline{\tau},p'}}\right) \ge C_{\overline{\tau},p'}(\delta_{\mathcal{X}})\left(\frac{\tau}{C_{\overline{\tau},p'}}\right).$$

Finally, note that $C_{\overline{\tau},p'} \ge \frac{p'-1}{2}C_{\overline{\tau}}^{-1}$, with $C_{\overline{\tau}}$ from item 2 of Lemma 3.2 and thus $C_{\overline{\tau},p'}\rho_{\mathcal{X}^*}(\overline{\tau})/\overline{\tau} \ge \frac{p'-1}{2}\overline{\tau}$.

*Claim* 4: By assumption we have $\delta^{j_p(x)}_{\mathcal{F},x}(\tau) \ge \phi(\tau)$ for $\tau \le \overline{\tau}$ and by Proposition 3.7 we have for $\tau > \overline{\tau}$ $\delta^{j_p(x)}_{\mathcal{F},x}(\tau) \ge \tau\delta^{j_p(x)}_{\mathcal{F},x}(\overline{\tau})/\overline{\tau}$ and thus $\delta^{j_p(x)}_{\mathcal{F},x}(\tau) \ge \tilde{\phi}(\tau)$ with

$$\tilde{\phi}(\tau) := \begin{cases} \phi(\tau), & \tau \le \overline{\tau}, \\ \tau\phi(\overline{\tau})/\overline{\tau}, & \tau > \overline{\tau}. \end{cases}$$

So by Proposition 3.9 we have for all $x^* \in S_{\mathcal{X}^*}$

$$\rho^{j_p^*(x^*)}_{\mathcal{F}^*,x^*}(\tau) = \left(\delta^{x^*}_{\mathcal{F},j_p^*(x^*)}\right)^*(\tau) \le \tilde{\phi}^*(\tau).$$

Now just observe that for $\tau < \phi(\overline{\tau})/\overline{\tau}$ we have

$$\frac{\tilde{\phi}^*(\tau)}{\tau} = \sup_{0\le t}\left[t - \frac{\tilde{\phi}(t)}{\tau}\right] = \sup_{0\le t\le\overline{\tau}}\left[t - \frac{\phi(t)}{\tau}\right] \to 0, \quad \tau \to 0,$$

as $\phi$ is nondecreasing. So by part 2 of the theorem we see that $\mathcal{X}^*$ is uniformly smooth from which it follows that $\mathcal{X}$ is uniformly convex [9, Prop. 1.e.2]. □

*Remark* 4.2 One can see from the above proof that in the asymptotic case $\overline{\tau} \to 0$ one can choose the constant $C_{\overline{\tau},p}$ such that

$$C_{\overline{\tau},p} \to \begin{cases} 2, & \mathcal{X} \text{ is not 2-smooth}, \\ 1+p, & \mathcal{X} \text{ is 2-smooth}. \end{cases}$$

These constants are not sharp for every space $\mathcal{X}$, but at least in the asymptotic case the constants are much simpler than the ones given in [5]. For best known constants with respect to $L^p$ spaces we refer to [13] and [14].

The above theorem combined with Proposition 3.6 gives us upper and lower bounds on the Bregman divergence for $\|x-y\| \le \overline{\tau}\|x\|$. However, as for large $\|x-y\|$ the Bregman divergence will be dominated by the term $\frac{1}{p}\|y\|^p$ it is not difficult to also find bounds that hold for all $x, y \in X$. Further one can additionally conclude bounds for the symmetric Bregman divergence,

$$\Delta^{\text{sym}}_{\mathcal{F}}(x,y) := \Delta^{j_p(x)}_{\mathcal{F}}(y,x) + \Delta^{j_p(y)}_{\mathcal{F}}(x,y) = \langle j_p(x) - j_p(y), x-y\rangle,$$

from our theorem. These two claims are shown in the following two propositions.

**Proposition 4.3** *For some fixed $p > 1$ let $\mathcal{F} = \frac{1}{p}\|\cdot\|^p$ and let $\phi\colon \mathbb{R}^+ \to \mathbb{R}^+$ be nondecreasing. Let $V = \mathcal{X} \setminus \{0\} \times \mathcal{X}$ and define the statements*:

$$\exists C, c > 0\ \forall (x,y) \in V, \|x-y\| \le c\|x\|\colon \quad \Delta_{\mathcal{F}}^{j_p(x)}(y,x) \le C\|x\|^p \phi\left(\frac{\|x-y\|}{\|x\|}\right), \tag{a}$$

$$\exists C > 0\ \forall (x,y) \in V\colon \quad \Delta_{\mathcal{F}}^{\mathrm{sym}}(x,y) \le C \max\bigl\{\|x\|, \|y\|\bigr\}^p \phi\left(\frac{2\|x-y\|}{\max\{\|x\|, \|y\|\}}\right), \tag{b}$$

$$\exists C > 0\ \forall (x,y) \in V\colon \quad \Delta_{\mathcal{F}}^{j_p(x)}(y,x) \le C \max\bigl\{\|x\|, \|y\|\bigr\}^p \phi\left(\frac{2\|x-y\|}{\max\{\|x\|, \|y\|\}}\right). \tag{c}$$

*Then* (a) $\Rightarrow$ (b) $\Rightarrow$ (c). *Obviously one also has* (c) $\Rightarrow$ {(a) *with $\phi$ replaced by $\phi(2\cdot)$*}.

*Proof* We only show that (a) implies (b) as (b) $\Rightarrow$ (c) follows trivially. Without loss of generality let $c \le 1$. First of all assume $\frac{\|x-y\|}{\|x\|} > c$. Then by

$$\frac{\|x-y\|}{\|x\|}\frac{\|x\|}{\|y\|} = \frac{\|x-y\|}{\|y\|} \ge \frac{\|y\| - \|x\|}{\|y\|} \ge 1 - \frac{\|x\|}{\|y\|}$$

one can see that regardless of whether we have $\|x\|/\|y\| > 1/2$ or $\|x\|/\|y\| \le 1/2$ one always has $\frac{2\|x-y\|}{\|y\|} > c$. So by

$$\begin{aligned}\Delta_{\mathcal{F}}^{\mathrm{sym}}(x,y) &= \bigl\langle j_p(x) - j_p(y), x-y\bigr\rangle \le \|x\|^p + \|y\|^p + \|x\|^{p-1}\|y\| + \|y\|^{p-1}\|x\| \\ &\le 4\max\bigl\{\|x\|, \|y\|\bigr\}^p\end{aligned}$$

we find that

$$\Delta_{\mathcal{F}}^{\mathrm{sym}}(x,y) \le \frac{4}{\phi(c)} \max\bigl\{\|x\|, \|y\|\bigr\}^p \phi\left(\frac{2\|x-y\|}{\max\{\|x\|, \|y\|\}}\right).$$

Now consider the case $\|x-y\|/\|x\| \le c \le 1$. In this range we can conclude (b) from (a) as $\|y\| \le 2\|x\|$, so that

$$\phi\left(\frac{\|x-y\|}{\|x\|}\right) \le \phi\left(\frac{2\|x-y\|}{\|y\|}\right). \qquad \square$$

**Proposition 4.4** *For some fixed $p > 1$ let $\mathcal{F} = \frac{1}{p}\|\cdot\|^p$, let $\phi\colon \mathbb{R}^+ \to \mathbb{R}^+$ be nondecreasing and $\phi(\tau) > 0$ for $\tau > 0$. Let $V = \mathcal{X} \setminus \{0\} \times \mathcal{X}$ and define the statements*:

$$\exists C, c > 0\ \forall (x,y) \in V, \|x-y\| \le c\|x\|\colon \quad \Delta_{\mathcal{F}}^{j_p(x)}(y,x) \ge C\|x\|^p \phi\left(\frac{\|x-y\|}{\|x\|}\right), \tag{d}$$

$$\exists C > 0\ \forall (x,y) \in V\colon \quad \Delta_{\mathcal{F}}^{j_p(x)}(y,x) \ge C \max\bigl\{\|x\|, \|y\|\bigr\}^p \phi\left(\frac{\|x-y\|}{\max\{\|x\|, \|y\|\}}\right), \tag{e}$$

$$\exists C > 0\ \forall (x,y) \in V\colon \quad \Delta_{\mathcal{F}}^{\mathrm{sym}}(x,y) \ge C \max\bigl\{\|x\|, \|y\|\bigr\}^p \phi\left(\frac{\|x-y\|}{\max\{\|x\|, \|y\|\}}\right). \tag{f}$$

*Then* (d) $\Rightarrow$ (e) $\Rightarrow$ (f).

*Proof* The proof is very similar to the previous proof so we just sketch it. We look at three different cases. By Proposition 3.7 we know that $\delta_{\mathcal{F},x}^{j_p(x)}$ is nondecreasing, so (d) gives also for $\|x-y\|/\|x\| \geq c$ $\Delta_{\mathcal{F}}^{j_p(x)}(y,x) \geq C\|x\|^p\phi(c)$ and thus

$$\Delta_{\mathcal{F}}^{j_p(x)}(y,x) \geq \begin{cases} C\|x\|^p\phi(\frac{\|x-y\|}{\|x\|}), & \frac{\|x-y\|}{\|x\|} \leq c, \\ C\|x\|^p\phi(c), & c \leq \frac{\|x-y\|}{\|x\|} < N, \\ C_{p,\phi,N}\|y\|^p\phi(\frac{\|x-y\|}{\|y\|}), & N \leq \frac{\|x-y\|}{\|x\|}, \end{cases}$$

for sufficiently large $N > 3$, where the last line follows from the fact that $\frac{\|x-y\|}{\|x\|} \geq N$ implies $\|y\| \geq (N-1)\|x\|$, which implies $\frac{\|x-y\|}{\|y\|} \geq 1-(N-1)^{-1} \geq 1/2$ and thus it suffices to see that the Bregman divergence will be dominated by $\|y\|^p$, for sufficiently large $N$. To conclude (e) one then basically has to redefine the constants. Equation (f) follows trivially. □

To conclude this chapter we combine the results and summarize the most important inequalities.

**Corollary 4.5** *Let $\mathcal{X}$ be a Banach space and $\mathcal{F}(x) = \frac{1}{p}\|x\|^p$ for $p > 1$. Then there exist constants $C_1, C_2 > 0$ such that for all $x, y \in \mathcal{X}$ we have*

$$\Delta_{\mathcal{F}}^{j_p(x)}(y,x) \leq C_1 \max\bigl\{\|x\|, \|y\|\bigr\}^p \rho_{\mathcal{X}}\left(\frac{2\|x-y\|}{\max\{\|x\|, \|y\|\}}\right) \tag{7}$$

*and*

$$\Delta_{\mathcal{F}}^{j_p(x)}(y,x) \geq C_2 \max\bigl\{\|x\|, \|y\|\bigr\}^p \delta_{\mathcal{X}}\left(\frac{\|x-y\|}{3\max\{\|x\|, \|y\|\}}\right). \tag{8}$$

*If the space $\mathcal{X}$ is s-smooth, then there exists $C > 0$ and for all $\overline{\tau} > 0$ also $C_{\overline{\tau}} > 0$ such that*

$$\Delta_{\mathcal{F}}^{j_p(x)}(y,x) \leq \begin{cases} C\|x-y\|^s, & p = s, \\ C_{\overline{\tau}}\|x\|^{p-s}\|x-y\|^s, \quad \textit{for } \frac{\|x-y\|}{\|x\|} \leq \overline{\tau}, & p \neq s. \end{cases} \tag{9}$$

*If the space $\mathcal{X}$ is r-convex, then there exists $\tilde{C} > 0$ and for all $\overline{\tau} > 0$ also $\tilde{C}_{\overline{\tau}} > 0$ such that*

$$\Delta_{\mathcal{F}}^{j_p(x)}(y,x) \geq \begin{cases} \tilde{C}\|x-y\|^r, & p = r, \\ \tilde{C}_{\overline{\tau}}\|x\|^{p-r}\|x-y\|^r, \quad \textit{for } \frac{\|x-y\|}{\|x\|} \leq \overline{\tau}, & p \neq r. \end{cases} \tag{10}$$

*Proof* By Proposition 3.6 we have

$$\|x\|^p \delta_{\mathcal{F},x/\|x\|}^{j_p(x)/\|x\|^{p-1}}\left(\frac{\|x-y\|}{\|x\|}\right) \leq \Delta_{\mathcal{F}}^{j_p(x)}(y,x) \leq \|x\|^p \rho_{\mathcal{F},x/\|x\|}^{j_p(x)/\|x\|^{p-1}}\left(\frac{\|x-y\|}{\|x\|}\right).$$

Thus item 1 of Theorem 4.1 and the $s$-smoothness show the bound (9) for $x \in \mathcal{X}$ and $y$ such that $\|x-y\| \leq \overline{\tau}\|x\|$. Similarly item 3 of Theorem 4.1 and the $r$-convexity show (10) for $x \in \mathcal{X}$ and $y$ such that $\|x-y\| \leq \frac{p'-1}{2}\overline{\tau}\|x\|$. As this holds true for all $\overline{\tau} > 0$ one can just replace $\tilde{\overline{\tau}} = \frac{2\overline{\tau}}{p'-1}$. Apply Proposition 4.3 and Proposition 4.4 to conclude from this the uniform bounds for all $x, y \in \mathcal{X}$. □

**Acknowledgements**
I thank my supervisor Thorsten Hohage for many helpful comments. Financial support by Deutsche Forschungsgemeinschaft through grant CRC 755, project C09, and RTG 2088 is gratefully acknowledged.

**Funding**
Financial support given by Deutsche Forschungsgemeinschaft through grants CRC 755 and RTG 2088.

**Availability of data and materials**
Not applicable.

**Competing interests**
The author declares that he has no competing interests.

**Authors' contributions**
There is only one author. The author read and approved the final manuscript.

## Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 15 August 2018 Accepted: 20 December 2018 Published online: 08 January 2019

**References**
1. Bregman, L.M.: The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. USSR Comput. Math. Math. Phys. **7**(3), 200–217 (1967)
2. Burger, M.: Bregman distances in inverse problems and partial differential equations. In: Advances in Mathematical Modeling, Optimization and Optimal Control. Springer Optimization and Its Applications, pp. 3–33. Springer, Cham (2016)
3. Censor, Y., Zenios, S.A.: Parallel Optimization: Theory, Algorithms, and Applications. Numerical Mathematics and Scientific Computation. Oxford University Press, New York (1997)
4. Reem, D., Reich, S., Pierro, A.D.: Re-examination of Bregman functions and new properties of their divergences. Optimization, 1–70 (2018)
5. Xu, Z.-B., Roach, G.F.: Characteristic inequalities of uniformly convex and uniformly smooth Banach spaces. J. Math. Anal. Appl. **157**(1), 189–210 (1991)
6. Zălinescu, C.: On uniformly convex functions. J. Math. Anal. Appl. **95**(2), 344–374 (1983)
7. Azé, D., Penot, J.-P.: Uniformly convex and uniformly smooth convex functions. Ann. Fac. Sci. Toulouse Math. **4**(4), 705–730 (1995)
8. Asplund, E.: Positivity of duality mappings. Bull. Am. Math. Soc. **73**(2), 200–203 (1967)
9. Lindenstrauss, J., Tzafriri, L.: Classical Banach Spaces II: Function Spaces. Ergebnisse der Mathematik und Ihrer Grenzgebiete, vol. 97. Springer, Berlin (1979)
10. Butnariu, D., Censor, Y., Reich, S.: Iterative averaging of entropic projections for solving stochastic convex feasibility problems. Comput. Optim. Appl. **8**(1), 21–39 (1997)
11. Butnariu, D., Reich, S., Zaslavski, A.J.: There are many totally convex functions. J. Convex Anal. **13**(3–4), 623 (2006)
12. Borwein, J., Guirao, A.J., Hájek, P., Vanderwerff, J.: Uniformly convex functions on Banach spaces. Proc. Am. Math. Soc. **137**(3), 1081–1091 (2008)
13. Xu, Z.-B.: Characteristic inequalities of $L^p$ spaces and their applications. Acta Math. Sin. **32**, 209–218 (1989) (Chinese)
14. Xu, Z.-B., Zhang, Z.-S.: Another set of characteristic inequalities of $L^p$ spaces. Acta Math. Sin. **37**, 433–439 (1989) (Chinese)

**Submit your manuscript to a SpringerOpen journal and benefit from:**

- ► Convenient online submission
- ► Rigorous peer review
- ► Open access: articles freely available online
- ► High visibility within the field
- ► Retaining the copyright to your article

**Submit your next manuscript at ► springeropen.com**